RESEARCH REPORT
(for RTOPs and Grants)

1. Title Directed Exploration of Complex Systems		**2. Date Prepared** 09 29 2008
3. Performing Organization: Jet Propulsion Laboratory		**4. RTOP/Grant No.**
4.A. JPL Project Number: **102294-982745.03.10**	**(Per GSK Policy, this serves as the Work Authorization Document)**	**4.C. NASA WBS NUMBER** **982745.02.02.03.14**
5. Investigator **Telephone** Michael C. Burl 818-393-1791	**6. NASA Program Manager** **Joseph Bredekamp**	**7. NASA Division** Science Mission Directorate
8. Reference **NRA Number: NNH07ZDA001N-AISR Other:**		

9. Funding Profile:	**FY'08 Prior Approvals**	**FY'09 Current Guideline**	**FY'09 Current Request**	**FY'09 Current Overguide**	**FY'10 Next Request**
	$140.0 K	$140.0 K	$ 140.0 K	$ 0.0 K	$ 0.0 K

10. Description

FY'08 Technical Accomplishments:

1. Established subcontract with SwRI.
2. Hired Esther Wang, who will be a senior in the Dept. of Applied and Computational Mathematics at Caltech this year, under the Summer Undergraduate Research Fellowship (SURF) program.
3. SwRI Co-I's have completed an initial set of 864 grid runs that simulate asteroid collisions and the subsequent gravitational interactions of the fragments. The grid trials sample a 5-dimensional parameter space, which includes new compositional parameters that enable the target body and or impacting body to be modeled as solid rock, a rubble pile, or a mixture of the two. The scientific question being addressed in these simulations particular is to understand which input parameters will result in an asteroid family with fragment size distributions similar to the Karin family. These runs have already yielded some interesting scientific insights.
4. A grading script was developed to measure the agreement between the raw simulation output and the observed Karin distribution. Thresholding the grading script output provides a binary-valued label for each simulation run.
5. Through an informal collaboration with Caltech (A. Holub, P. Perona), which grew out of a presentation by the PI describing the Directed Exploration task, we have developed a new active learning algorithm paradigm based on reducing the uncertainty (entropy) of the likely labels of the unlabeled points. A method for using a committee of classifiers to deduce probabilistic label assignments was another innovative aspect of the approach. This algorithm was applied to efficiently learn object recognition models for a large set of object categories. The paper describing this work was awarded "Best Paper" at the CVPR Workshop on Online Learning for Classification.
6. Adapted the entropy-based active learning algorithm to use a Gaussian Process (GP) predictive model along with an expected information gain criteria to sequentially choose the most informative point to run through the simulator. Since the GP model directly provides probabilities, this modified approach requires significantly less computation than the committee of classifiers.
7. Delivered code for the algorithm to SwRI for initial tests with the asteroid simulation data.

FY'08 Publications and Presentations:

1. A. Holub, M.C. Burl, P. Perona, "Entropy-based Active Learning for Object Recognition", CVPR Workshop on Online Learning for Classification, (Jun 2008) – *Best Paper Award*
2. M.C. Burl, "Directed Exploration of Complex Systems", NASA Applied Information Systems Research (AISR) PI Meeting, College Park, MD, (May 2008)
3. E. Wang, M.C. Burl, "Learning Simplified Predictive Models of Complex Dynamical Systems", NASA Conference on Intelligent Data Understanding (CIDU), (Sep 2008)

FY'09 Plans:

1. Convert the active learning algorithm from Matlab to C to improve the runtime and increase the size of the datasets that can be handled. Focus additional research on algorithmic modifications to improve the scalability.
2. Submit technical paper on latest approach to the SIAM Data Mining Conference.
3. Develop solutions for the situation in which multiple simulation trials can be started simultaneously (or trails started before other trials are completed) as this situation is very important in practice (multiple simulation trials in progress concurrently across a cluster).
4. Systematically, evaluate the new approach against other active learning algorithms.
5. Evaluate algorithms "in the loop" with the actual asteroid collision simulator.

Approval:	**Date:**	**Concurrence:**	**Date:**